"Express Mail" Mailing Label No. **EV352469945US**

PATENT APPLICATION
ATTORNEY DOCKET NO. BPA02-0001

METHOD AND APPARATUS FOR USING FACETED METADATA TO NAVIGATE THROUGH INFORMATION RESOURCES

Inventors: Bradley P. Allen, Cormac Twomey and John B. Jensen

Related Application

[0001] This application hereby incorporates by reference and claims priority under 35 U.S.C. §119 to U.S. Provisional Patent Application No. 60/406,959, filed on 28 August 2002, entitled "Teapot Server," by inventor Bradley P. Allen (Attorney Docket No. BPA02-0001PSP).

BACKGROUND

Field of the Invention

[0002] The present invention relates to systems for searching through collections of information resources. More specifically, the present invention relates to a method and an apparatus for using faceted metadata to facilitate navigation through information resources.

Related Art

[0003] Traditional free-text searching or keyword indexing is a simple and inexpensive search solution used by many web sites. Unfortunately, text searching requires users to guess at keywords. Typical users under-specify their searches and are disappointed in the results. To improve the precision of the search, the user's only option is to guess again with a new combination of keywords. However, text search engines give the user little or no guidance on how to improve the search results.

[0004] When text search isn't sufficient, many web designers fall back on hierarchical navigation. This can be seen in department-store e-commerce web sites. In such systems, the searcher must navigate through a pre-defined hierarchy, making a series of one-from-many selections before reaching a list of products that might or might not be appropriate.

[0005] Although hierarchical navigation works well for simple domains and for well-informed searchers, it acts as a barrier when the desired product doesn't fit the hierarchy, or the searcher does not share the web designer's view of classification. As soon as a searcher confronts a one-from-many choice he doesn't know how to make, the search fails. Often the searcher hits a dead end and is forced to backtrack with no guarantee that the desired object can be found at all.

[0006] The Resource Description Framework (RDF) has been developed to describe characteristics of web pages and other resources. RDF can improve web searching by making metadata about web resources explicit. This allows systems use of metadata to perform search queries for the associated web resources.

[0007] Existing approaches to expressing RDF queries are generally based on either the syntax and semantics of SQL, or of logic programming environments

like Prolog. These existing approaches typically present results in a format, such as RDF, that is usable by machines, but is not easily decipherable by humans. Moreover, these existing approaches do not have built-in support for query reformulation and refinement during a search session, and they do not gracefully handle large result sets.

[0008] Hence, what is needed is a method and an apparatus for performing web site searching without the above-described problems.

SUMMARY

[0009] One embodiment of the present invention provides a system that uses faceted metadata to facilitate navigation through information resources. During operation, the system receives a query from a client at a server. The system then performs the query on metadata, wherein the metadata contains facets that describe characteristics of the information resources. While performing the query, the system generates results that identify information resources that satisfy the query. Next, the system constructs a response containing the results, the query, and suggestions on how to refine the query. The system then sends the response to the client, thereby allowing the client to refine the query. In this way, the client and server can work together in a stateless manner to refine the query without having to maintain state information about the query on the server.

[0010] In a variation on this embodiment, the suggestions on how to refine the query include suggested values for facets of the metadata. These suggested values can include frequently occurring values for facets of the metadata.

[0011] In a variation on this embodiment, the suggestions can include instructions on how to display the suggestions to a user.

[0012] In a variation on this embodiment, upon receiving the response from the server at the client, the system displays the results and the suggestions on

how to refine the query to a user associated with the client. Upon subsequently, receiving a command from the user to modify the query, the system modifies the query in accordance with the command to generate a new query, and then sends the new query from the client to the server.

[0013] In a variation on this embodiment, modifying the query in accordance with the command can involve: using one of the suggestions to define a new query term; defining a new query term that is not associated with one of the suggestions (such as a new text search query term); and removing a query term from the query.

[0014] In a variation on this embodiment, displaying the results and the suggestions further involves displaying a representation of the state of the query to the user.

[0015] In a variation on this embodiment, the query contains a specification of facets to be used in organizing the results, and conditions that results must satisfy.

[0016] In a variation on this embodiment, the system automatically creates an initial query by: scanning through facets of the metadata; generating suggestions for facets that have commonly occurring values; and allowing a user to select one or more of the suggestions to create the initial query.

[0017] In a variation on this embodiment, the suggestions on how to refine the query can additionally specify frequencies for commonly occurring values of facets in the metadata.

[0018] In a variation on this embodiment, prior to receiving the query, the system initializes a database containing the metadata by: receiving the metadata in Resource Description Framework (RDF) format; and storing the metadata in the database.

[0019] In a variation on this embodiment, the query and the response are encoded in eXtensible Markup Language (XML) documents that are transferred between the client and the server.

[0020] In a variation on this embodiment, the format of the query and the response are specified by a query language that facilitates navigation using faceted metadata.

BRIEF DESCRIPTION OF THE FIGURES

[0021] FIG. 1 illustrates a computer system in accordance with an embodiment of the present invention.

[0022] FIG. 2 presents a flow chart illustrating initialization operations in accordance with an embodiment of the present invention.

[0023] FIG. 3 presents a flow chart illustrating the process of creating an initial query in accordance with an embodiment of the present invention.

[0024] FIG. 4 presents a flow chart illustrating how a query is processed in accordance with an embodiment of the present invention.

[0025] FIG. 5A presents a page for an exemplary system in accordance with an embodiment of the present invention.

[0026] FIG. 5B presents another page for the exemplary system in accordance with an embodiment of the present invention.

[0027] FIG. 5C presents yet another page for the exemplary system in accordance with an embodiment of the present invention.

[0028] FIG. 5D presents another page for the exemplary system in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

[0029] The following description is presented to enable any person skilled in the art to make and use the invention, and is provided in the context of a particular application and its requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present invention. Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

[0030] The data structures and code described in this detailed description are typically stored on a computer readable storage medium, which may be any device or medium that can store code and/or data for use by a computer system. This includes, but is not limited to, magnetic and optical storage devices such as disk drives, magnetic tape, CDs (compact discs) and DVDs (digital versatile discs or digital video discs), and computer instruction signals embodied in a transmission medium (with or without a carrier wave upon which the signals are modulated). For example, the transmission medium may include a communications network, such as the Internet.

Computer System

[0031] FIG. 1 illustrates a computer system 100 in accordance with an embodiment of the present invention. Computer system 100 includes a metadata search server 102, a web server 105 and a web client 112. These entities are coupled together by a network (not shown), which can generally include any type of wire or wireless communication channel capable of coupling together computing nodes. This includes, but is not limited to, a local area network, a wide

area network, or a combination of networks. In one embodiment of the present invention, the network includes the Internet.

[0032] Web client 112 can generally include any node on the network including computational capability and including a mechanism for making service requests across the network. In FIG. 1, web client 112 is associated with a user 114, who runs applications on web client 112.

[0033] Web server 105 can generally include any computational node including a mechanism for servicing requests from a client for computational and/or data storage resources. Web server 105 generally services requests from web client 112. Note that web server 105 is itself a client of metadata search server 102. More specifically, a navigation application 106 on web server 105 interacts with metadata search server 102.

[0034] Metadata search server 102 uses faceted metadata to facilitate navigation through information resources in accordance with an embodiment of the present invention. More specifically, metadata search server 102 performs searches and related navigational operations involving data stored within database 110. To this end, metadata search server 102 includes a query engine 103 that makes queries on metadata contained within database 110, and a navigation engine 104 facilitates navigational operations involving the metadata.

[0035] During operation, web server 105 submits a query document 107 to metadata search server 102. In response to query document 107, metadata search server 102 returns a response document 108. Response document 108 contains enough information to allow web server 105 to refine the query without having to maintain state information about the query on metadata search server 102. These query-processing operations are described in more detail below with reference to FIGs. 2-5.

System Initialization

[0036] FIG. 2 presents a flow chart illustrating initialization operations in accordance with an embodiment of the present invention. The system starts by initially creating an empty "model" for storing metadata (step 202). Next, the system loads metadata into the model, for example from an RDF file. In one embodiment of the presenting invention, this involves defining an RDF feed and its properties (step 204), and then using the feed to load an RDF file into the model (step 206).

[0037] Next, the system can generate an initial query (step 208). Referring to FIG. 3, generating an initial query can involve scanning through facets of the metadata (step 302), and then generating suggestions for facets with commonly occurring values (step 304). The system can then allow the user to select one or more suggestions or perform other modifications on the query to create an initial query (step 306).

Query Processing

[0038] FIG. 4 presents a flow chart illustrating how a query is processed in accordance with an embodiment of the present invention. First, the system defines a query (step 402). In doing so, the system can automatically create an initial query as outlined above, or alternatively, the user can explicitly create a query. The client then submits the query to the server (step 404).

[0039] Note that a query can include simple query terms that specify conditions such as "attribute = value", wherein the attribute is an attribute of the resource (also referred to as a facet) and the value is a value for the attribute. Additionally, the "=" relation can be replaced with other relations, such as, "<", ">", "≤" or "contains". For example, the simple query term "attribute contains string" can be used to perform a text search for the string

within the attribute. These simple query terms can additionally be compounded using operators, such as, NOT, OR, and AND. (Note that the AND operator is typically implicit.)

[0040] The query can also suggest particular facets to be used in organizing the results, and can also include instructions on how to display the particular facets.

[0041] Next, the server performs the query to generate a set of results that satisfy the query conditions (step 406). The server then constructs a response containing the results, the query and suggestions on how to refine the query (step 408). For example, the suggestions can include frequently occurring values (or other interesting values) for facets of the metadata. As mentioned above, the suggestions can also include suggestions on how to display the suggestions to a user. At this point, the server can forget state information relating to the query, which means that the server will be "stateless" when it receives the next query.

[0042] The server then returns the response to the client (step 410). The client then displays the query result, the query state and the suggestions to the user (step 412). The user can then perform a number of actions to refine the query. For example, the user can select one of the suggestions, can define a new query term, can define a new text search term, or can remove a term from the query. These user actions generate one or more commands.

[0043] Upon receiving these commands from the user, the client refines the query and submits it to the server (step 414). The system then returns to step 406 to process the new query. This process can be repeated as long as the user wants to continue refining the query.

[0044] In a variation on this embodiment, the format of the query and the response are specified in an XML-based query language that facilitates navigation using faceted metadata.

Example

[0045] FIG. 5A presents an initial page of an exemplary search in accordance with an embodiment of the present invention. This initial page indexes 1239 articles from medical journal articles. This page allows the user to perform a number of actions. By entering a text string into the field in the upper-right portion of the screen, a user can perform a keyword search, as many web sites do.

[0046] Alternatively, the user can select values for facets that describe properties, such as "subject", "author", "publisher", and "date". Selecting a value narrows the search. For example, selecting the subject "Antibiotics" narrows the collection down to 28 matching articles as is illustrated in FIG. 5B. Note that metadata (and possible other information) associated with the matching articles appears on the right-hand side of FIG. 5B.

[0047] Further refinement suggestions appear on the left-hand side of FIG. 5B. In additional to facilitating navigation, these refinement suggestions provide useful information on the frequency of suggestions. For example, from FIG. 5B we can see that 25 of the 28 articles containing the subject "Antibiotics" also contain the subject "Cystic Fibrosis".

[0048] The query state appears on the top of the left-hand side of FIG. 5B. In this example, the query state indicates that the query has the condition "subject 'Antibiotics'".

[0049] The user can further refine this query, for example, by selecting the publisher "J-Infect-Dis" which is an abbreviation for the "Journal of Infectious Diseases". This narrows the collection down to 3 articles as is illustrated in FIG. 5C.

[0050] Note that the selected criteria can be removed by deselecting the checkboxes next to the conditions in the query state (which appears in the upper

left-hand side of FIG. 5C). For example, deselecting the "subject 'Antibiotics'" checkbox broadens the collection to 14 articles as is illustrated in FIG. 5D.

[0051] As can be seen from the example above, the present invention helps the user to quickly navigate to the right answer. The user does not have to guess at keywords because the system provides the available keywords. Moreover, the user is not locked into a rigid hierarchy because the user can partition the collection in any way the user wants to. During a search, the user can always see a few results that provide an intuitive guide to the success of the search. If the user hits a dead end, the user is able to back out by relaxing previous selections.

[0052] The foregoing descriptions of embodiments of the present invention have been presented for purposes of illustration and description only. They are not intended to be exhaustive or to limit the present invention to the forms disclosed. Accordingly, many modifications and variations will be apparent to practitioners skilled in the art. Additionally, the above disclosure is not intended to limit the present invention. The scope of the present invention is defined by the appended claims.